INRIA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P 105
78153 Le Chesnay Cedex
France
Tél. 954 90 20

Rapports de Recherche

N° 77

PLANAR CIRCUIT COMPLEXITY AND THE PERFORMANCE OF VLSI ALGORITHMS

John E. SAVAGE

Mai 1981

PLANAR CIRCUIT COMPLEXITY

and the

PERFORMANCE OF VLSI ALGORITHMS*

by

*John E. SAVAGE***

Department of Computer Science
Brown University
Providence,RI

January, 1981
Revised April, 1981

---

* The research reported here was supported in part by NSF grant MCS 76-20023 by the University of Paris-Sud and by INRIA, Rocquencourt, France.

** Currently on sabbatical leave at the University of Paris-Sud, Orsay and INRIA, Rocquencourt, France.

ABSTRACT

The size of the smallest planar circuit for a problem, its planar complexity, is shown to provide a lower bound on the performance of VLSI algorithms as measured by the products AT^2 and A^2T where A is the VLSI chip area and T is the number of cycles executed. These bounds apply to a simple model of VLSI chips. The planar complexity of several problems is considered and it is shown that integer binary division, certain reciprocals and powers all have planar complexities quadratic in the lengths of their inputs. The same results hold for predicates associated with these problems.

The relative strengths of the two inequalities showing that A^2T and AT^2 are both bounded below by planar complexity is considered. It is shown that problems exist with linear planar complexity for which AT^2 is the better measure and others for which A^2T is superior. Finally the area required when inputs that are normally grouped together, such as components in binary numbers, are contiguous on the boundary of a chip is considered. It is shown that problems exist for which the area needed becomes square in the lengths of the inputs, while in the absence of the condition, linear area can be acheived.

RESUME

Nous montrons que la taille du plus petit circuit planair pour un problème, donc sa complexité planaire, donne une borne inférieure sur la performance des algorithms VLSI mesurés par les produits AT^2 et A^2T, ou A est la surface de la puce VLSI et T est le nombre de cycles qui sont exécutés. Ces bornes sont établies pour un modèle très simple des puces VLSI. La complexité planaire de quelques problèmes est étudiée et nous montrons que la division des entiers binaires, certains inverses et puissances ont des complexité planaires qui sont quadratiques en le nombre des données. Les mêmes résultats sont obtenus pour des prédicats qui sont associés avec ces problèmes.

Nous considérons la précision relative des deux inégalités données par AT^2 ou A^2T. Il est montré qu'il existe des problèmes avec une complexité planaire qui est linéaire et pour lesquelles AT^2 est la meilleure mesure et d'autres pour lesquelles A^2T est supérieur. Nous considérons aussi la surface qui est requise quand les données qui sont normalement groupées, comme les composantes des entiers binaires, sont contigües sur la frontière d'une puce. Il est montré qu'il y a des problèmes pour lesquelles la surface devient quadratique en le nombre de données quand cette condition est vérifiée mais pour lesquelles la surface devient linéaire autrement.

1 - INTRODUCTION

In 1979 Thompson [1] demonstrated that, under a suitable model for VLSI chips, the product AT^2 of chip area A and time T to compute the Fast Fourier Transform (FFT) on n inputs must satisfy $AT^2 = \Omega(n^2)$. This model accounts for the consumption of chip area by wires. This work was subsequently extended to include sorting [2], and bounds on this problem and the FFT were improved to $\Omega(n^2 \log^2 n)$. Brent and Kung [3] introduced a somewhat different model for VLSI chips in which the area occupied by wires and circuit elements is convex. They demonstrate that $AT^2 = \Omega(n^2)$ to multiply two n-bit integers, a result obtained independently with the original Thompson model by Abelson and Andreae [4]. Brent and Kung show that $A = \Omega(n)$. They also present algorithms that come close to meeting their lower bounds. Savage [5] obtained bounds of $AT^2 = \Omega(p^4)$ with bath models for $p \times p$ matrix multiplication, inversion and transitive closure. Algorithms previously given for matrix multiplication and transitive closure by Kung and Leiserson [6] and Guibas et al. [7] were shown to be optimal. Preparata and Vuillemin [8] subsenquently introduced a family of optimal matrix multiplication algorithms for $\Omega(\log n) \leq T \leq O(n)$.

Vuillemin [9] has extended the Brent-Kung model to pipelined chips. If P is the period of computations, that is, the time between the production of two successive results, he shows that $AP^2 = \Omega(n^2)$ for transitive functions on n inputs. He also demonstrates that $A = \Omega(n)$ for these problems. Yao [10] considers VLSI algorithms for $x + y * z$, over a finite field F, as well as predicates and derives bounds of the form $AT^2 = \Omega(n^2)$. Similarly results are obtained by Lipton and Sedgewick [18] for a number of predicates. They also consider the role of randomness and nondeterminancy.

In this paper we present a new framework for the consideration of VLSI algorithms. Our model as presented in Section 2.1 is a simplification of the Thompson and Brent-Kung models. It is basically a sequential machine realized from discrete components. We show in Section 2.3 that the area, A, of the chip, with wire widths and spacings of at least λ, and the number of cycles, T, needed to compute a function $f : \{0,1\}^n \rightarrow \{0,1\}^m$, must satisfy the three following computation inequalities :

$$C(f) \leq \nu(A/\lambda^2)T, \quad C_p^s(f) \leq 12\nu^2 \min[(A/\lambda^2)T^2, (A/\lambda^2)^2 T]$$

Here ν is the number of layers in the chip, $C(f)$ is the size of the smallest circuit for f and $C_p^s(f)$ is the size of its smallest planar circuit in which there is one node associated with each input variable of f. The derivation of these inequalities is preceded in Section 2.2 by a discussion of some of the properties of these two circuit measures. There it is shown that planar circuit size is at worst quadratic in the unrestricted circuit size measure.

The Planar Separator Theorem of Lipton and Tarjan [11] is reformulated in Section 3 to provide conditions on single-output and multiple-output functions from which lower bounds on planar circuit size can be obtained. Lipton and Tarjan [11] have shown that the shifting function and matrix multiplication, as well as any functions that reduce to them through assignments to some input variables, have planar circuit sizes quadratic in the lengths of their inputs. Our new framework thus provides alternate proofs to previously obtained results. We prove that many new problems reduce to the shifting function. These include the e th power of a binary integer, $e > 1$, e th power of reciprocals, $e > 0$ for $e = g/2^k$, $g > 0$ and binary integer division. We also extend a proof of Vuillemin [9] to show that transitive functions have quadratic planar circuit size and improve the quadratic lower bound of Lipton and Tarjan [11] for matrix multiplication by about six orders of magnitude.

The area required by a chip is examined in Section 4. There we extend a method of Yao [10] which is used to obtain lower bounds to the area required by 1-output functions.

Section 5 contains a discussion ot the relative strengths and weaknesses of the three computational inequalities cited above. By considering binary sorting, which is shown to have linear planar circuit size and to require chip are which is logarithmic in the length of the input,

2 - COMPUTATIONAL INEQUALITIES FOR VLSI CHIPS

In this section, we introduce a model of VLSI circuits which is somewhat more general than those presented by Thompson [1] and Brent-Kung [3] and use it as a basis for deriving computational inequalities.

2.1 - The VLSI Model

The model is described as follows :

A1. The chip realises a sequential machine constructed from discrete boolean logic elements and straight wire segments.

A2. Wires have a width of λ and a separation and a length of at least λ. Each logic element occupies an area of at least λ^2. The chip has ν planes each of which may contain wires or logic elements.

A3. Inputs are read and outputs are produced at times that are data-independent.

A4. Each input variable is supplied exactly once to the chip.

Concerning assumption A1, we note that every sequential machine can be realized from discrete logic elements and flip-flops, and the latter can in turn be realized from logic elements using feedback. In practice, chips are realized at the level of transistors but designed at the level of flip-flops and gates. In the sequential machines the flip-flops need not all change state at the same time ; however, for all values of inputs the sequencing must be correct. We include logic circuits in the category of clocked sequential machines ; they execute one cycle of computation.

Assumption A2 reflects the fact that photographic resolution limits the width, length and separation of units and that the technology of chip fabrication limits the number of layers of material which can be deposited on a chip. Our assumption places the same lower bound, λ^2, on

we demonstrate that the inequality $A^2T = \Omega(C_p^s(f))$ is sometimes stronger than the inequality $A^2T = \Omega(C_p^s(f))$. On the other hand, another function is given with linear planar circuit size for which the reverse is true. Thus, care must be taken in ascribing undue importance to the AT^2 measure. This section is closed by a discussion of the importance of natural constraints on chip size. It is shown that if binary numbers are supplied to the boundary of an adder on a convex chip, while honoring the natural contiguity between components in these numbers, then area quadratic in the lengths of inputs is necessary. Linear area suffices if this constraint is dropped. Conclusions are stated in Section 6.

the area of any unit of a chip, although the area of units, such as logic elements and ports, do vary depending upon function. This only strengthens the lower bounds. Permitting several logic elements to be placed above one another on different planes runs contrary to current practice, but again only strengthens the lower bounds.

The third assumption presumes that the chip is to be used for a special purpose, say as part of a machine complex, so that the irregular reading or writing could cause a deterioration in system performance.

A chip that satisfies assumption A4, namely, it reads each input variable exactly once, is said to be semelective (This is a neologism formed from the latin words "semel", meaning once, and "lectio" meaning to read). Some of the computational inequalities that we derive do not depend on this assumption. However, the condition is apparently essential for the derivation of strong lower bounds on the chip area required for certain problems.

In the Thompson model [1] assumption A1 is modified by requiring that wires run on a rectangular grid while the type of processing elements, that is, whether Boolean gates or more complicated operations, perhaps with memory, is ignored. In [2] assumption A3 is dropped and A4 is augmented by requiring that each input and output variable be associated with unique nodes on the chip.

The Brent-Kung model [3] replaces assumption A1 by the assumption that the several layers of the chip describe an outer boundary that is convex. Otherwise, the remaining assumptions prevail, although A3 can be relaxed.

2.2 - Circuit complexity

A logic circuit is the graph of a Boolean straight-line algorithm (SLA). An SLA with Boolean inputs $x_1, x_2, \ldots, x_n$ is a sequence of steps $\beta_1, \beta_2, \ldots, \beta_k$ in which

1. $\beta_i \in \{x_1, x_2, \ldots, x_n, 0, 1\}$

or 2. $\beta_i = (h, \beta_a, \beta_b)$, $a, b < i$,

h is 2-input, 1-output Boolean function and $x_1, x_2, \ldots, x_n$ are indeterminates to which values in {0,1} may be assigned. A step is said to be of type 1 or 2 according to its classification above. For each step β_i of an SLA, a function $Res(\beta_i)$ is defined recursively as follows :

1. If β_i is of type 1, $Res(\beta_i) = \beta_i$

2. If β_i is of type 2, $Res(\beta_i) = h(Res(\beta_a), Res(\beta_b))$

An SLA computes $f = (f_1, \ldots, f_m)$, $f_j : \{0,1\}^n \to \{0,1\}$, if for each j there is a step β_{j_i} such that $Res(\beta_{j_i}) = f_j$.

From each SLA a graph called a logic circuit may be derived. Each step is associated with a node and the node associated with β_i of type 2 (a logic element) has edges directed into it from the nodes associated with β_a and β_b. The condition $a, b < i$ insures that this graph has no directed loops.

We will be interested in SLA's for which the corresponding graph is planar and those for which it is not. We will also be interested in SLA's which are semelective and those which are not. The semelective condition will affect one of our complexity measures, planar combinational complexity, but not the other.

Definition 1

A Boolean function $f : \{0,1\}^n \to \{0,1\}^m$ has combinational complexity C(f) if the minimal number of steps of type 2 in an SLA for f is C(f). Such a function has a planar combinational complexity $C_p(f)$ if it is the minimal number of steps of type 2 in an SLA for f whose logic circuit is planar. If in addition, the class of SLA's considered is semelective, the corresponding measures are called $C^s(f)$, $C_p^s(f)$. □

Observe immediately that $C(f) = C^s(f)$ since any non-semelective circuit can be converted into a semelective one without the addition of logic elements ; only edges need be added. When it is required that the circuit be planar, the relationship between $C_p(f)$ and $C_p^s(f)$ could be as large as quadratic, although no examples are known which illustrate this.

Standard combinational complexity has been extensively studied (see Savage [12]) but as yet no lower bounds have been derived that are non-linear in the number of inputs and outputs except for functions defined essentially by diagonalization. However, the planar combinational complexity of several multiple-output functions is known to be at least quadratic (see Lipton and Tarjan [11]). The following theorem establishes some simple relationships between $C(f)$, $C_p(f)$ and $C_p^s(f)$.

THEOREM 1 : *For all functions* $f : \{0,1\}^n \to \{0,1\}^m$

$$C(f) \le C_p(f) \le C_p^s(f) \le 6[C(f)]^2$$

PROOF : The first two inequalities are immediately obvious. The third follows from the observation that a (potentially non-planar) circuit with C logic elements has 2C edges, one for each input to a logic element, and the number pairs of crossing edges thus cannot exceed $\binom{2C}{2}$ or $C(2C-1)$. Each pair of crossing edges can be reduced to a planar circuit with 3 EXCLUSIVE OR's, as shown in Figure 1. Thus, from an arbitrary circuit of C elements a planar circuit can be constructed using at most $C + 3C(2C-1) \le 6C^2$ elements. □

Some of the functions on n inputs discussed later will be seen to have $C(f) = O(n(\log n)^2)$ and $C_p(f) = \Omega(n^2)$ thus demonstrating that the spread suggested by these inequalities can be nearly ach ved. However, at this time no function is known for which $C_p^s(f) = \Theta([C(f)]^2)$.

The following result demonstrates that although individual functions may have a nearly quadratic gap between their planar and non-planar circuit sizes, for all almost all functions $f : \{0,1\}^n \to \{0,1\}^m$, the gap is at worst a factor of n.

THEOREM 2 : *For any* $0<\delta<1$ *a fraction of at least* $1-2^{-\delta m 2^n}$ *of the functions* $f : \{0,1\}^n \to \{0,1\}^m$ *have*

$$C(f) \geq \frac{m2^n}{n}(1-\delta-O(n))$$

if $\log_2 m = O(n)$, *and for all such functions and all values of* m *and* n,

$$C_p^s(f) \leq 7\ m\ 2^n$$

PROOF : A proof is given here of the upper bound. The lower bound is obtained by a standard counting argument which is given in the Appendix.

The bound is obtained by constructing a planar semelective circuit for an arbitrary function $f : \{0,1\}^n \to \{0,1\}^m$. This is done by first realizing a planar decoder circuit, a circuit with inputs $x_1, x_2, \ldots, x_n$ and 2^n outputs, each associated with one of the 2^n different minterms in these variables. As shown in Figure 2, a decoder $f_d^{(n)}$ on n inputs can be realized from a decoder $f_d^{(n-1)}$ on n-1 inputs using 2^n logic elements plus 3 elements for each crossing of the input line x_n and the 2^{n-1} outputs of the circuit for $f_d^{(n-1)}$. Thus,

$$c_p^s(f_d^{(n)}) \leq c_p^s(f_d^{(n-1)}) + 5\ 2^{n-1}$$

logic elements. To realize $f : \{0,1\}^n \to \{0,1\}^m$, supply inputs $x_1, \ldots, x_n$ to m decoders, as shown in Figure 3, using at most $n(n-1)/2$ crossings of input edges for each decoder. Therefore,

$$C_p^s(f) \leq m(5\ 2^n + 2^{n-1} + 3n(n-1)/2) \leq 7\ m2^n$$

which completes the proof. □

For standard combinational complexity Lupanov [13] has shown that the lower bound that applies to almost all functions can be achieved up to a small constant (see also Savage [12], page 116). The exact value of $C_p^s(f)$ for $f : \{0,1\}^n \to \{0,1\}^m$ for n large and $\log_2 m = O(n)$ is not known and its determination remains an open problem.

Before going on to evaluate the planar combinational complexity of individual functions, we demonstrate its role in the analysis of the performance of VLSI circuits by deriving computational inequalities based upon it.

2.3 - Computational inequalities

Three sets of computational inequalities are now derived that apply to the use of area and time by VLSI chips. The derivations constitute simulations of chips by planar and non-planar circuits.

THEOREM 3 : *Let* $f : \{0,1\}^n \to \{0,1\}^m$ *be computed in* T *cycles by a VLSI chip that has area* A. *Then, the inequality*

$$C(f) \leq \nu \frac{A}{\lambda^2} T$$

must be satisfied.

PROOF : This is a restatement of an earlier result by Savage [14]. It is obtained by constructing a circuit for f using at most the indicated number of logic elements from copies of the chip. A stack of T chips is formed and where loops exist on the chip in flip-flops, the loops are broken and signals fed to the corresponding position on the chip above. A logic circuit of at most $(\nu A/\lambda^2)T$ logic elements is created that computes the same function computed by the chip. Thus, C(f) cannot exceed this value. □

It is important to note that T in this theorem is the number of cycles executed and not the time in seconds to compute f. The physics of chips may in fact require that the duration of a cycle be dependent on the geometry of the chip, as suggested by Seitz [15] and Chazelle and Monier [16]. These observations also apply to the following theorem.

THEOREM 4 : *Let* $f : \{0,1\}^n \to \{0,1\}^m$ *be computed in* T *cycles by a VLSI chip of area* A. *Then, the following inequality*

$$C_p^s(f) \leq 12v^2 \min((A/\lambda^2)T^2, (A/\lambda^2)^2 T)$$

must be satisfied. If the chip algorithm is not semelective, the inequality holds when $C_p^s(f)$ *is replaced by* $C_p(f)$.

PROOF : A stack of chips with loops broken and reconnected as described in the proof of the preceding theorem, is used here also. However, a planar circuit is now constructed. This is done by first transforming the chip itself so that individual logic elements are visible when viewed from above, so that no wires are seen to overlap and so that at most one pair of wires cross at a point. Then, the stack of modified chips is transformed into a planar circuit. This is done in two ways to produce the two inequalities. In the first case, the copies of the chip in the stack are given incremental shifts, as suggested in Figure 4a, in order to make all wire crossings visible. Crossings are now made planar using the schema of Figure 1 and a planar circuit results. The second way in which a planar circuit is produced is that the T chips are placed side by side, as suggested in Figure 4b, and crossings on the chip and those created by the wires connecting them are replaced to produce a planar circuit. If the VLSI algorithm is semelective, the same will hold true for the circuits so constructed. The details of the proof follow.

Let n_1 and w be the number of logic elements and straight-wire segments on the chip. Let n_{CN} be the number of places at which exactly two wire segments meet. (Wires can meet on the same plane or can meet between adjacent planes. We assume without loss of generality that the angle between the wires is different from 0° and 180°).

When seen from above the chip, wires on different planes may appear to cross. Let n_{CR} be the number of such places exclusive of those places where exactly two wires meet. Then,

$$\frac{n_L}{\nu} + n_{CR} + n_{CN} \leq A/\lambda^2 \qquad (1)$$

since, except for logic elements, the places at which wires meet or cross occupy disjoint regions of the surface of the chip, each of area at least λ^2, when seen from above, and at most ν logic elements, each of area at least λ^2, overlap.

Now make all logic elements infinitesimal in size and all wires infinitesimal in width. At places where more than two wires cross, as illustrated in Figure 5, give them an infinitesimal displacement so that only pairs of wires are seen to cross. If the wires that cross consist of straight segments, at most $\nu(\nu-1)/2$ crossings will become visible. If wires cross at points of connection, at most 2ν wires are involved and at most $2\nu^2$ crossings can become visible. If n_{CR}^v is the number of visible crossings of pairs of wires on the displaced chip, then

$$n_{CR}^v \leq 2\nu^2 \, n_{CR} \qquad (2)$$

Also, let n_{CN}^v be the number of visible points of connection on the chip after displacement. Since a crossing could contain ν points of connection, we have

$$n_{CN}^v \leq n_{CN} + \nu n_{CR} \qquad (3)$$

From (1), (2) and (3) we have

$$n_L + n_{CR}^v + n_{CN}^v \leq 3\nu^2 \, A/\lambda^2 \qquad (4)$$

which must hold. The number of wires on the chip, w, cannot exceed

$$w \leq \nu A/\lambda^2 \qquad (5)$$

since each must occupy an area of at least λ^2 and they can reside on at most ν planes.

When a stack of loop-free chips is displaced as shown in Figure 4a, each pair of crossing wires on the original chip generates T^2 crossings, as indicated in Figure 6a. Similarly a pair of connecting wires, if shifted in the right direction, will generate $T(T-1)/2$ crossings (see Fig. 6b). It follows that the graph produced by the displacement of Figure 4a will have $n_L T$ logic elements and at most $(n_{CR}^v + n_{CN}^v)T^2$ crossings. Replacing each of these by 3 logic elements we have from (4) a planar circuit with at most $9(\nu/\lambda)^2 AT^2$ logic elements for f, which produces the first of the two inequalities.

To establish the connections of Figure 4b, wires are run in parallel between two copies of the chip. There are at most w of these wires and each can intersect each of the w original wires at most once. Thus, at most w^2 crossings are created for each pair of chips. The graph thus produced has $n_L T$ logic elements and at most $(n_{CR}^v + n_{CN}^v)T + w^2(T-1)$ crossings. Replacing crossings by 3 logic elements we construct a planar circuit which from (4) and (5) has at most

$$9(\nu/\lambda)^2 AT + 3(\nu/\lambda^2)^2 A^2(T-1) \leq 12\nu^2(A/\lambda^2)^2 T$$

logic elements. This establishes the second inequality. □

Of these two inequalities, the one involving $(A/\lambda^2)T^2$ is stronger when $A/\lambda^2 \geq T$. This is a condition that is usually satisfied for functions for which quadratic lower bounds to AT^2 have been derived. This includes transitive functions and matrix multiplication. The inequality $A/\lambda^2 \geq T$ is satisfied primarily because such problems require a large surface area, as will be seen below. Later, problems will be presented for which the required surface area is small in which case the inequality involving $(A/\lambda^2)^2 T$ is the stronger.

As stated in theorem 4, if the VLSI algorithms are semelective, that is, if each input variable is read exactly once, then each circuit constructed in the proof has exactly one node associated with each variable. If the chip algorithm is not semelective, neither are the circuits. The semelective condition has been essential in all previous derivations of quadratic lower bounds to AT^2.

Recently Valiant [24] has noted that the lower bounds on the space-time product for uni-processor machines operating on straight-line algorithms that are derived using the Grigoryev formulation [25] translate directly into lower bounds on the product A^2T. This has value since the lower bounds derived by this method are obtained without the semelective constraint. We can immediately translate previous results to show that A^2T must be quadratic in the number of input variables for n-degree polynomial multiplication over GF(2) [25], the discrete Fourier transform , sorting and merging [26], binary integer multiplication [27], and matrix inversion [28]. Also, Gregoryev has shown a p^3 lower bound for p×p matrix multiplication [25].

3 - APPLICATION OF THE PLANAR SEPARATOR THEOREM

Lipton and Tarjan [11] show that planar vertex N-node graphs have separators of at most $2\sqrt{2N}$ nodes that divide the graph into two disconnected subgraphs of about equal size, as stated below in the Planar Separator Theorem (PST).

THEOREM 5 [11] : *Let* G *be any* N*-vertex planar graph with non-negative vertex costs summing to no more than 1. Then, the vertices of* G *can be partitioned into three sets* U, V, W *such that no edge joins a vertex in* U *with a vertex in* V, *neither* U *nor* V *has vertex cost exceeding 2/3, and* W *contains no more than* $2\sqrt{2}\sqrt{N}$ *vertices.*

Lipton and Tarjan [11] have shown that semelective planar circuits which contain an n-superconcentrator, which realize $p \times p$ Boolean matrix multiplication where $n = p^2$ or which compute the shifting function $f_s^{(n)}$ must have a number of logic elements which is $\Omega(n^2)$. We immediately conclude from Theorem 4 that

$$AT^2 = \Omega(n^2)$$

for semelective algorithms for each of these problems. These results have been previously demonstrated by direct means using the Thompson and Brent-Kung models. The superconcentrator result has not been explicitly stated but is immediately obvious following the analysis of Thompson [1, 2] or that of Brent and Kung [3]. The latter obtain the result for the shifting function and apply it to the problem of binary integer multiplication. Savage [5] has obtained the stated type of bound for matrix multiplication but the coefficient on n^2 is much larger. The Lipton-Tarjan bound is $c\ n^2$ for $c = 8 \times 10^{-10}$.

Of considerable interest is whether or not similar bounds can be obtained for one-output functions. This issue arises because the method of proof used by authors of each of the above results consists of measuring

information flow from inputs to a comparable number of outputs. Yao [17] has a framework for consideration of information flow between two processors that compute a Boolean function. He has used this formulation together with the Brent-Kung model to obtain lower bounds to AT^2 for several 1-output problems on graphs [10], including testing for isomorphism of two n-vertex graphs. Yao's technique for measuring information flow is to consider a partition of the variables of a function into two sets U and V, with U and V of about equal size. He then considers the table of values of f where rows and columns are indexed by the tuples of values of variables in U and V, respectively. In this table, a mono-chromatic rectangle is the intersection of rows and columns, each intersection of which contains the same element, 0 or 1. If d(f) is the minimum number of non-overlapping such rectangles needed to cover the table, then he demonstrates that at least $\lceil \log_2 d(f) - 2 \rceil$ bits must be exchanged between processors. If the table contains a square $M \times M$ subtable with a single 1 (or 0) in each row and column, clearly $d(f) \geq \log_2 M$. Lipton and Sedgewick [18] have restated this condition on Boolean functions and have independently used it to obtain good lower bounds to AT^2 for 1-output functions. This condition is presented below using our terminology.

Definition 2 :

A function $h : \{0,1\}^s \rightarrow \{0,1\}^t$ is a subfunction of $f : \{0,1\}^n \rightarrow \{0,1\}^m$ if it can be obtained by suppressing some output variables and by an assignment to a subset of the input variables. Also, h is a subfunction of f with respect to J if $J \subseteq \{1,2,\ldots,n\}$ and h is obtained from f by an assignment of values to the variables x_j, $j \in J$, of f.

□

Definition 3 :

A function $f : \{0,1\}^n \rightarrow \{0,1\}$ is w-separated if there exists a subset X of its variables such that for any partition of X into two sets A, B with $|A|, |B| \leq 2|X|/3$, there exist at least 2^w pairs $\{(\underline{a}_i, \underline{b}_i)\}$ of variables such that $f(\underline{a}_i, \underline{b}_j) = 1$ if $i = j$ and 0 otherwise. A function $f : \{0,1\}^n \rightarrow \{0,1\}^m$ is also w-separated if it contains a subfunction which is w-separated.

□

The following theorem is an analog of the Yao result and has been stated by Lipton and Sedgewick for a variant of the Brent-Kung model. It illustrates the importance of the above definition.

THEOREM 6 : If $f : \{0,1\}^n \to \{0,1\}^m$ *is w-separated, then*

$$C_p^s(f) \geq w^2/32$$

PROOF : Since a lower bound is to be derived, we may assume without loss of generality that $f : \{0,1\}^n \to \{0,1\}$ and is itself w-separated. Consider a semelective planar circuit for f. Apply the PST and let U, V, W constitute a partition of the graph when each input in X has cost $1/|X|$ and all other nodes have cost 0. Then, U and V each contains at most $2|X|/3$ inputs. Let the respective sets of inputs in U and V be A and B as well as a partition of inputs in W so that $|A|, |B| \leq 2|X|/3$.

Consider the subgraph of the circuit with nodes W. It has at most $2|W|$ nodes directed in from U and V. Thus, given 2^w sets of distinct pairs of values for input variables $\{(\underline{a}_i, \underline{b}_i)\}$ corresponding to A and B, by the pigeon-hole principle, if $2|W| \leq w-1$, there must exist some two pairs $(\underline{a}_r, \underline{b}_r)$ and $(\underline{a}_s, \underline{b}_s)$ for which the inputs to this subgraph are the same. Of course, outputs of W are also the same. Without loss of generality, let the node of the planar graph that is associated with f reside in U or W. By the preceding argument, it follows that inputs $(\underline{a}_r, \underline{b}_r)$ and $(\underline{a}_r, \underline{b}_s)$ provide the same values for inputs and outputs to W, and hence the value of f is the same on inputs $(\underline{a}_r, \underline{b}_r)$ and $(\underline{a}_r, \underline{b}_s)$. But this contradicts the definition of f, so

$$4\sqrt{2N} \geq 2|W| \geq w$$

which is the desired conclusion. □

Definition 4 :

Consider a function $f : \{0,1\}^n \to \{0,1\}^m$ and let X and Y be subsets of the sets of input and output variables of f, respectively. It is said to have a w-flow if for all partitions of X U Y into two sets A, B with $|A|, |B| \leq 2(|X| + |Y|)/3$, there are sets $S \subset X$ and $S^* \subset Y$, where S is a subset of A and S^* a subset of B or vice versa, such that for some assignment to variables not in S, the resulting subfunction h of f from S to S^* has at least 2^w points in the image of its domain. A function f also has a w-flow if it contains a subfunction with this property. □

The following theorem illustrates the importance of such functions. It should be noted that all previous quadratic lower bounds to AT^2 or to planar complexity are obtained by means of lower bounds on some type of information flow.

THEOREM 7 : If $f : \{0,1\}^n \to \{0,1\}^m$ *has a w-flow, then*

$$C_p^s(f) \geq w^2/8$$

PROOF : Without loss of generality let f itself be its subfunction which has a w-flow. Consider a semelective planar graph or circuit for f with N nodes. Apply the PST by giving each input vertex in X and output vertex in Y cost equal to $1/(|X| + |Y|)$. Let U, V, W be the partition of the nodes of this graph guaranteed by the PST. Then, $|W| \leq 2\sqrt{2N}$ and U and V each have cost $\leq 2/3$ or they contain sets A^* and B^*, respectively, of inputs and outputs where $|A^*|, |B^*| \leq 2(|X| + |Y|)/3$. Also, if W contains p input nodes and q output nodes, then $|A^*| + |B^*| + p + q = n + m$. Distribute the nodes in W between A^* and B^* to form sets A and B where $|A|, |B| \leq 2(|X| + |Y|)/3 = 2(n + m)/3$.

Since f has a w-flow there are sets S, S^* of inputs and outputs where, without loss of generality, $S \subset A \cap X$, $S^* \subset B \cap Y$. Then, for some assignment to variables not in S, the mapping induced from S to S^* by the resulting subfunction of f has at least 2^w points in the image of its domain. However, if $|W| \leq w-1$, the outputs of the subgraph with nodes in W can assume at most $2^{|W|}$ different values which is less than 2^w. Since all paths from inputs in $S \subset U \cup W$ to outputs $S^* \subset V \cup W$ must pass through W, it follows that f cannot be computed if $|W| \leq w-1$. Thus, $2\sqrt{2N} \geq |W| \geq w$ from which the desired result follows. □

Functions which have a w-flow also have corresponding predicates, as given in definition 4, which are w-separated, as shown below.

Definition 5 :

Given a function $f : \{0,1\}^n \rightarrow \{0,1\}^m$, $f(x_1,\ldots,x_n) = (y_1,\ldots,y_m)$, its associated predicate $F : \{0,1\}^{n+m} \rightarrow \{0,1\}$ is defined by

$$F(x_1,\ldots,x_n,z_1,\ldots,z_m) = \begin{cases} 1 & f(x_1,\ldots,x_n) = (z_1,\ldots,z_m) \\ 0 & f(x_1,\ldots,x_n) \neq (z_1,\ldots,z_m) \end{cases}$$

where $x_1,\ldots,x_n$ and $z_1,\ldots,z_m$ are the variables of F. □

THEOREM 8 : *If* $f : \{0,1\}^n \rightarrow \{0,1\}^m$ *has a w-flow, then its associated predicate* $F : \{0,1\}^{n+m} \rightarrow \{0,1\}$ *is w-separated.*

PROOF : Assume without loss of generality that f is its own subfunction which has a w-flow. Let X and Y be the subsets of inputs and outputs with respect to which f has this property.

Let Q be the subset of the inputs of F corresponding to X and Y. Consider an arbitrary partition of Q into sets A, B where $|A|, |B| \leq 2(|Q|)/3$. This generates a partition of the union of X and Y of inputs and outputs of f, respectively.

By the w-flow condition of f, there exist sets $S \subset X$, $S^* \subset Y$ such that $S \subset A$ and $S^* \subset B$ or vice-versa and such that by some assignment to variables not in S, f defines a subfunction from S to S^* with at least 2^w points in the image of its domain. Let $\{\underline{X}_i\}$ be assignments of f to inputs which yield the set $\{f(\underline{X}_i)\}$ of at least 2^w distinct values over Y. If follows that if the auxiliary variables Z of F are given values $\underline{Z}_i = f(\underline{X}_i)$, then the condition for F to be w-separated is met. □

As a consequence of Theorem 11, any function $f : \{0,1\}^n \rightarrow \{0,1\}^m$ that has a large w-flow immediately has an associated predicate for which a large lower bound to its semelective planar combinational complexity can be derived. Thus, almost automatically we exhibit a large family of 1-output functions for which strong lower bounds to AT^2 and A^2T exist.

Bounds are now derived on the w-flow for two important classes of problems, transitive functions and Boolean matrix multiplication, as well as functions which contain these as subfunctions. Of the proofs presented, that for transitive functions is an extension to a proof given by Vuillemin [9] for the Brent-Kung model. The matrix multiplication result borrows some ideas from a proof by Savage [5] for the Thompson and Brent-Kung models.

We now state the Lipton-Tarjan lower bound to planar combinational complexity for the shifting function as a bound on w-flow.

THEOREM 9 : The shifting function $f_s^{(n)} : \{0,1\}^{n+k} \rightarrow \{0,1\}^{2n-1}$ *has a* n/18-*flow.*

This same result holds if the 2's complement of the value of $f_s^{(n)}$ is taken since this operation realizes a bijective mapping.

Definition 6 [9]

A Boolean function $h_G(x_1,\ldots,x_p,s_1,\ldots,s_k) = (y_1,\ldots,y_p)$ computes a permutation group G if for each permutation $g \in G$ there exist values for $s_1,\ldots,s_k$ such that $y_i = x_{g(i)}$, $1 \le i \le p$. Such a function is transitive if in addition for each $1 \le i, j \le p$ there exists $g \in G$ such that $g(i) = j$. In general, a function $f_G : \{0,1\}^n \to \{0,1\}^m$ is transitive of degree p if f_G has a subfunction $h_G(x_1,\ldots,x_p,s_1,\ldots,s_k)$, which is transitive. □

The function $f_{cs}^{(n)}(x_1,\ldots,x_n,s_1,\ldots,s_k) = (y_1,\ldots,y_n)$ which performs all cyclic shifts of $x_1,\ldots,x_n$ is transitive of degree n as is the function ABC defined by the product of three $\sqrt{n} \times \sqrt{n}$ matrices A, B, C with addition and multiplication over GF(2)[9]. The integer multiplication function defined by the product of two n-bit binary numbers with the result in binary, denoted $f_m^{(n)} : \{0,1\}^{2n} \to \{0,1\}^{2n}$ contains as a subfunction the shifting function $f_s^{(n)}(x_0,\ldots,x_{n-1},s_1,\ldots,s_k) = (y_0,y_1,\ldots,y_{2n-2})$ where by assignment to the control variables $s_1,\ldots,s_k$ the mappings

$$y_j = \begin{cases} x_{j-t} & t \le j \le t+n-1 \\ 0 & \text{otherwise} \end{cases}$$

for each $0 \le t \le n-1$. However, the shifting function is not transitive.

LEMMA 1 : Let a_1, a_2, b_1, b_2 *be integers satisfying the following constraints :*

$$0 \le a_1, a_2, b_1, b_2 \le m, \quad a_1 + b_1 = m, \quad a_2 + b_2 = m$$

$$2m/3 \le a_1 + a_2 \le 4m/3, \quad 2m/3 \le b_1 + b_2 \le 4m/3$$

for some integer m. Then,

$$P = \frac{1}{m} \max\,(a_1 b_2,\ a_2 b_1) \ge 2m/9$$

PROOF : Let $c_1 = a_1 + a_2$, $c_2 = b_1 + b_2$. In a_1b_2 use the following substitutions :

$$b_2 = c_2 - b_1, \qquad b_1 = m - a_1$$

to give

$$T_1 = a_1b_2 = a_1c_2 - ma_1 + a_1^2$$

Similarly substitue

$$a_2 = c_1 - a_1, \qquad b_1 = m - a_1$$

to give

$$T_2 = a_2b_1 = c_1m - c_1a_1 - a_1m + a_1^2$$

It follows that

$$T_1 \geq T_2$$

if and only if

$$a_1 \geq c_1/2$$

since $c_1 + c_2 = 2m$. Thus, P is smallest when $a_1 = c_1/2$. But this implies that $a_1 = a_2$, $b_1 = b_2 = c_2/2$. Consequently,

$$P \geq c_1c_2/4m$$

where $c_1 + c_2 = 2m$ and $2m/3 \leq c_1, c_2 \leq 4m/3$.

Under these constraints, c_1c_2 is a convex function of c_1 with a minimum at $c_1 = 2m/3$ and $c_1 = 4m/3$, from which the desired conclusion follows. □

THEOREM 10 : Let $f_G^{(n)} : \{0,1\}^n \to \{0,1\}^n$ *be a transitive function. Then it has a* $2n/9$*-flow.*

PROOF : As Vuillemin [9] states, the sets $G_{ij} = \{g|g(i) = j\}$ all have the same size and $|G_{ij}| = |G|/n$.

In definition 4, let X and Y be the permuted inputs and outputs, respectively. Given a partition A, B of $X \cup Y$ with $|A|, |B| \leq 4n/3$, for each input in $A \cap X$ and output in $B \cap Y$, there are $|G|/n$ group elements that identify them. A similar statement applies to pairings between $B \cap X$ and $A \cap Y$. Summing over group elements $g \in G$ and inputs in X, the number of pairings is at least

$$Q = \frac{|G|}{n} \max(|A \cap X|\ |B \cap Y|, |B \cap X|\ |A \cap Y|)$$

Since each $g \in G$ is a permutation on inputs, there exists $g \in G$ that maps at least $P = Q/|G|$ inputs in S to outputs in S^*. Invoking lemma 1 with $m = n$, we have that a flow of $w \geq 2n/9$ exists. □

This result guarantees that $C_p^s(f_G^{(n)}) \geq n^2/162$. The proof given above is an extension of that given in [9] for the Brent-Kung model. In this model, one can assume that a common non-zero lower bound on $|A \cap Y|$, $|B \cap Y|$ exists, which is not true here.

We cite the following lemma whose proof is obtained by using the identity $a \oplus b = (a - b)^2$.

LEMMA 2 : *Let* $\underline{e} = (e_0, e_1, \ldots, e_{m-1})$ *and* $\underline{d} = (d_0, d_1, \ldots, d_{m-1})$ *be two Boolean vectors. If*

$$\sum_{i=0}^{m-1} e_i \oplus d_i \leq \rho m$$

then the weights of $\underline{e}$ *and* $\underline{d}$, $|e|$ *and* $|d|$, *satisfy*

$$||e| - |d|| \leq \rho m$$

THEOREM 11 : *The matrix multiplication function* $f_{MM}^{(p)}$ *defined by the product* $C = D \times E$ *of two* $p \times p$ *Boolean matrices has a* $p^2/12$*-flow.*

PROOF : As in the proof of the lower bound to AT^2 for the Thompson and Brent-Kung models for this problem [5], we note that if E is chosen to be a cyclic permutation matrix which shifts the columns of D k places for some $0 \leq k \leq p-1$, then

$$C_{ij} = d_{i,\, j+k}$$

for all $0 \leq i, j \leq p-1$, where subscript addition is modulo p. Similarly, if D is a cyclic permutation matrix, then for some $0 \leq l \leq p-1$,

$$C_{ij} = e_{i+l,\, j}$$

for all $0 \leq i, j \leq p-1$, where again subscript addition is modulo p.

In definition 3, let X be the components of D and E and Y the components of C. Given a partition A, B of $X \cup Y$ with $|A|, |B| \leq 2(|X| + |Y|)/3 = 2p^2$, at the risk of some confusion, we let $[d_{ij}]$, $[e_{ij}]$ and $[c_{ij}]$ take on the significance of characteristic matrices where a component has value 1 if the corresponding component of the corresponding matrix is in A and 0 otherwise.

From the preceding discussion, if

$$P_k = \sum_{i,j} c_{ij} \oplus d_{i,\, j+k}, \qquad Q_l = \sum_{i,j} c_{ij} \oplus e_{i+l,\, j}$$

and $P_k \geq \beta p^2$ or $Q_l \geq \beta p^2$ for some values of $0 \leq k, l \leq p-1$, then choosing E as the k-th cyclic permutation matrix or D as the l-th, a flow $w \geq \beta p^2/2$ can be achieved. If $\beta = 1/6$, we are done. Otherwise, assume that $P_k < \beta p^2$ and $Q_l < \beta p^2$ for all $0 \leq k, l \leq p-1$. We show that this leads to a contradiction.

Using the following inequality

$$a \oplus b = (a \oplus c) \oplus (b \oplus c) \leq (a \oplus c) + (b \oplus c)$$

we have

$$R_{k,l} = \sum_{i,j} d_{i,j+k} \oplus e_{i+l,j} \leq P_k + Q_l < 2\beta p^2$$

Now use $a \oplus b = (a - b)^2$ to expand $R_{k,l}$ and sum over $0 \leq k, l \leq p-1$ to give

$$p^2(d + e) - 2de < 2\beta p^4$$

where d and e are weights of $[d_{ij}]$ and $[e_{ij}]$, that is, the number of components of the two matrices which are in A. Thus, if $s = d + e$ it follows from the above inequality by maximizing de that

$$s(2p^2 - s) < 4\beta p^4$$

from which we conclude that either

$$s < p^2(1 - \sqrt{1-4\beta}) \quad \text{or} \quad s > p^2(1 + \sqrt{1-4\beta})$$

Also, using Lemma 2 and $P_k < \beta p^2$, $Q_l < \beta p^2$ we have

$$|c - d| < \beta p^2, \qquad |c - e| < \beta p^2$$

where c is the number of components of C in A. Then these two inequalities imply that

$$\frac{s}{2} - \beta p^2 < c < \frac{s}{2} + \beta p^2$$

Taking these inequalities together with the bounds on s, we have that $|A| = c + d + e$ must satisfy

$$|A| < \frac{3p^2}{2}(1 - \sqrt{1-4\beta}) + \beta p^2 \quad \text{or} \quad |A| > \frac{3p^2}{2}(1 + \sqrt{1-4\beta}) - \beta p^2$$

However, with $\beta = 1/6$ the constraints $|A|, |B| \leq 2p^2$ are violated. □

This proof borrows the idea of sets of cyclic permutation matrices to map inputs onto outputs from [5] but otherwise the proof is different. It also differs from the proof of [11] and gives a lower bound to planar combinational complexity that is about six orders of magnitude larger.

It is immediately obvious that the n-bit binary integer multiplication function $f_m^{(n)} : \{0,1\}^{2n} \to \{0,1\}^{2n}$ contains the shifting function $f_s^{(n)} : \{0,1\}^{n+k} \to \{0,1\}^{2n-1}$ as a subfunction, as does the Boolean convolution function $f_{BC}^{(n)}(x_1,\ldots,x_n,y_1,\ldots,y_n) = (z_2,\ldots,z_{2n})$ where $z_k = \sum_{i+j=k} x_i y_j$. It is also straightforward to reduce binary squaring $f_{SQ}^{(n)} : \{0,1\}^n \to \{0,1\}^{2n}$ to integer multiplication and then to shifting. The following result is more general and will be used to study a range of problems, including integer division, powers and inverse powers.

LEMMA 3 : *The function* $f_{SQ}^{(n,p)} : \{0,1\}^n \to \{0,1\}^m$, $m = 2n + \lceil \log_2 p \rceil$ *defined by* $f_{SQ}^{(n,p)}(X) = \lceil pX^2 \rceil$, *where* $p = q/2^k$, *integers* $0 \le k \le 2n/3$, $1 \le q \le 2^{\lfloor 2n/9 \rfloor}$, $0 \le X \le 2^n - 1$, *and where* X *and* $\lceil pX^2 \rceil$ *are represented in binary, contains* $f_s^{(\lfloor 2n/9 \rfloor - 3)}$ *as a subfunction for* $n \ge 15$.

PROOF : Let $X = a + b2^e$ for $e = \lfloor 2n/3 \rfloor$ and let $0 \le a \le 2^m - 1$, $0 \le b \le 2^l - 1$ for $m = \lfloor 2n/9 \rfloor$ and $l = \lfloor n/9 \rfloor$. Also, let $t = \lceil \log_2 q \rceil$, $t \le \lfloor 2n/9 \rfloor$ and let $k \le \lfloor 2n/3 \rfloor$. Then,

$$\lceil pX^2 \rceil = \lceil qa^2/2^k \rceil + qab2^{e+1-k} + qb^2 2^{2e-k}$$

since all terms but possibly the first are integers. Furthermore, the middle term appears in positions $e + 1 - k, \ldots, 2e - k - 1$ of the binary expansion of $\lceil pX^2 \rceil$ without overlap from the first and third terms. Thus, $\lceil pX^2 \rceil$ contains qab as a subfunction.

Now consider the following values for a :

$$\left\{ a = \left\lceil \frac{2^s}{q} \right\rceil \mid 1 \le a \le 2^m - 1 \right\}$$

This implies that $t \le s \le s_1$ where $s_1 \ge m + t - 2$. Since

$$\frac{u}{v} \le \left\lceil \frac{u}{v} \right\rceil \le \frac{u+v-1}{v}$$

for integers u, v ≥ 1, it follows that

$$2^s \le q\left\lceil \frac{2^s}{q} \right\rceil \le 2^s + q - 1$$

Hence

$$2^s b \le qab \le 2^s b + (q-1)b$$

for the chosen values of a. Then, since q satisfies

$$2^{t-1} < q \le 2^t$$

it follows that $(q-1)b < 2^{s_0}$ for $s_0 = t + 1$ so that positions $s_0, s_0 + 1, \ldots, s_1$ of qab contain $2^s b$. Here $s_1 - s_0 \ge m-1-2$.

It follows that qab contains as a subfunction, the function which shifts b by between s_0 and s_1 positions, or equivalently, by anywhere from 0 to $s_1 - s_0 \ge \lfloor n/9 \rfloor - 3$ positions. Thus the number of shifts is three less than the length of b. Thus, under these conditions, qab contains $f_s^{(l-3)}$, where $l = \lfloor n/9 \rfloor$.

THEOREM 12 : *The functions* $f_R^{(n,e)}$ *and* $f_p^{(n,e)}$ *defined by*

$$f_R^{(n,e)}(x,y) = \lceil (2^n/y)^e \rceil, \qquad f_p^{(n,e)} = \lceil 2^{2n} x^e \rceil$$

in which $1 \le x,y \le 2^n - 1$, $e = q/2^k > 0$, k *and* q *integers independent of* n, *and* $e > 1$ *for* $f_p^{(n,e)}$, *contain the shifting function* $f_s^{(m)}$, $m = e(n)$, *as a subfunction. For* $0 < e < 1$, $f_p^{(n,e)}$ *contains the 2's complement of the shifting function in* $e(n)$ *variables as a subfunction.*

PROOF : Consider $f_R^{(n,e)}$ and let $y = 2^a + z$. We use Taylor's expansion

$$\frac{1}{(1+\eta)^e} = 1 - e\eta + \frac{e(e+1)}{2}\eta^2 + R(\eta)$$

for $0 \le \eta$. It is straightforward to show that

$$|R(\eta)| \le \frac{e(e+1)(e+2)}{6}\eta^3$$

Thus,

$$\left\lceil \left(\frac{2^n}{2^a+z}\right)^e \right\rceil = \left\lceil 2^{(n-a)e} \cdot \frac{1}{(1+z/2^a)^e} \right\rceil$$

$$= 2^{(n-a)e} - e2^{(n-a)e-a} z + \frac{e(e+1)}{2} 2^{(n-a)e-2a} z^2$$

$$+ \left\lceil 2^{(n-a)e} R(z/2^a) \right\rceil$$

if $(n-a)e$ is an integer, $(n-a)e - 2a \ge 2k+1$ and $z < 2^{a/3}/(e+2)^{1/3}$. Also, in the binary expansion of this function all bits in $e(e+1)z^2$ appear unaltered in the output. Apply the preceding lemma with a the largest integer such $(n-a)e$ is an integer and $(n-a)e - 2a \ge 2k+1$. This implies that $a = e(n)$ which implies that this function contains the squaring function on $e(n)$ variables as a subfunction which in turn reduces to the shifting function with $e(n)$ variables.

To treat the function $2^{2n}x^e$, let $x = 2^a + z$ and use the following expansion :

$$(1 + \epsilon)^e = 1 + e\epsilon + \frac{e(e-1)\epsilon^2}{2} + R(\epsilon)$$

Here $|R(\epsilon)| \le e(e-1)\,|e-2|\,\epsilon^3/6$. An analysis similar to that given above produces the desired conclusion for $e > 1$. When $0 < e < 1$, it can also be shown that this function reduces to $2^p - q(2^k-q)z^2$ for p such that $2^p > q(2^k-q)z^2$ and $1 \le z \le 2^m-1$, $m = \lfloor (n-s)/3 \rfloor$, s a constant. Following Lemma 3, this reduces to the 2's complement of the shifting function. □

THEOREM 13 : *Let* $f_S^{(n)}$, $f_G^{(n)}$, $f_{MM}^{(p)}$, $f_{SQ}^{(n)}$, $f_D^{(n)}$, $f_P^{(n,e)}$ *and* $f_R^{(n,e)}$ *be the shifting function, a transitive function,* $p \times p$ *matrix multiplication, n-bit binary integer squaring, n-bit integer division, e th powers for* $e \neq 1$ *and e th inverse powers for* $e > 0$, *where* $e = q/2^k$. *Each of these functions and their associated predicates have semelective planar combinational complexities which are at least quadratic in the length of their inputs.*

Matrix inversion and the transitive closure of Boolean matrices are also of the type mentioned above since each can be reduced to Boolean matrix multiplication.

The bounds given above translate into bounds on AT^2 and A^2T. Yao [10] has quadratic lower bounds on AT^2 for several predicates including the Boolean function that tests for isomorphism of graphs. Lipton and Sedgewick [18] have the same type of bounds for several problems including recognition of palindromes, pattern matching and the predicate associated with binary integer multiplication. The latter result was obtained independently of that given above.

4 - BOUNDS ON CHIP AREA

We now consider the area required by semelective VLSI chips. The following class of functions requires large chip area, as shown later.

Definition 7 :

$P_{p,g}^{(n,m)}$ is the set of functions $f : \{0,1\}^n \to \{0,1\}^m$ for which there exist at least g distinct subfunctions of f with respect to J, for all sets J such that $|J| = p$.

This class has been studied in [12, p.34] for m = 1. Clearly $g \leq 2^p$.

Yao [10] derives a lower bound on the area required by a VLSI chip of the kind examined here. He considers the information that must be transfered in one direction between two processors with one way communication to compute a function f, denoted $C(I \to Z; f)$, and shows that the minimum of this quantity over all partitions of inputs into two sets of about equal size provides a lower bound to chip area. Our theorem stated below is a slight generalization of his result which will useful in Section 5. There we consider a function which is in $P_{p,g}^{(n,1)}$ for $g = 2^p$ and $p = \Theta(n/\log n)$ and thus requires large area. However, the function also has a small planar circuit size.

THEOREM 5 : *Let* $f : \{0,1\}^n \to \{0,1\}^m$ *in* $P_{p,g}^{(n,m)}$ *be computed by a semelective chip. Then, the chip area must satisfy*

$$A/\lambda^2 \geq \lceil \log_2 g \rceil /2$$

even if reading and writing by the chip is not done in a data-independent manner (condition A3).

PROOF : Let $s = \lceil \log_2 g \rceil$. Let k be the maximum number of inputs read in any one cycle by the chip. If $k \geq s/2$, the result follows immediately because at least k ports of area at least λ^2 each are required. If $k < s/2$, let r be the largest number of bits read by the chip over several cycles without exceeding p. Then $r \geq p-(k-1)$. Since $f \in P_{p,g}^{(n,m)}$, it is also in $P_{u,v}^{(n,m)}$ for $u = p-a$, $v = g2^{-a}$. It follows that if exactly p-a inputs are given to the chip and at most $\lceil \log_2 g \rceil -a-1$ flip-flops exist on the chip, the function f cannot be computed because these inputs determine at least $g2^{-a}$ distinct subfunctions yet the chip can be in at most $g2^{-a}-1$ states prior to receiving the remaining $n-(p-a)$ inputs. It follows that at least $\lceil \log_2 g \rceil -a$ flip-flops are required or A/λ^2 is at least this large. Replacing a by k-1 and noting that $k < s/2$ we have that in this case $A/\lambda^2 \geq s-(k-1) > s/2$, which establishes the desired result. We note that this argument holds whether or not inputs and outputs are read or produced in a data-independent manner. □

Several lower bounds have been obtained on chip area for multiple output functions realized by semelective chips. Brent and Kung [3] have a lower bound of $\Omega(n)$ for n-bit binary integer multiplication, Vuillemin [9] has shown that any transitive function of degree p requires area $A/\lambda^2 \geq p$, Baudet [19] has shown that the shifting function $f_s^{(n)} : \{0,1\}^{n+k} \to \{0,1\}^{2n-1}$ requires area $A/\lambda^2 \geq n/2$, and Heinz [20] has demonstrated that p x p matrix multiplication requires area linear in p^2.

5 - DISCUSSION

Much of the recent literature on the performance of VLSI algorithms concerns lower bounds to AT^2 which are quadratic in the lengths of inputs. These results reinforce the notion that this measure is basic and should be used to evaluate the performance of VLSI algorithms for all problems.

The two computational inequalities of Theorem 4 are restated below.

$$C_p^s(f) \leq 12v^2 \min[(A/\lambda^2)T^2, (A/\lambda^2)^2 T]$$

The $(A/\lambda^2)T^2$ inequality dominates when $A/\lambda^2 \geq T$. As shown in a preceding section, a large number of important problems have $C_p^s(f) = \Omega(n^2)$ for n the length of the input for f. For these same problems, the chip area required is linear in n. Thus, if the inequality given above is to be satisfied, then it must be that $T = O(\sqrt{n})$. Furthermore, it follows that $(A/\lambda^2)T^2$ is the better of the two bounds. These observations further reinforce the signifiance of this measure.

In this section we examine problems for which the $(A/\lambda^2)^2 T$ measure is superior to $(A/\lambda^2)T^2$. These are problems for which $C_p^s(f) = O(n)$. We also present a problem with linear planar combinational complexity that requires a large chip area, thus showing that $(A/\lambda^2)T^2$ is sometimes the better measure for simple functions. In addition, we consider the effect of the simple assumption that inputs that are normally grouped together, such as components of a binary number, are placed contiguously on a chip boundary. We show that this can require a chip area that is quadratic in the length of the input.

The sorting function $f_{SRT}^{(n,k)} : \{0,1\}^{nk} \rightarrow \{0,1\}^{nk}$ that sorts n k-bit integers $k \geq \lceil \log_2 n \rceil + 1$ can easily be shown to be transitive of degree n. Let the control inputs consist of the k-1 most significant bits in each integer. Ignore the corresponding bits in the output. By choosing the k-1 most significant bits to represent distinct (k-1)-tuples, the

function $f_{SRT}^{(n,k)}$ performs an arbitrary permutation of the n remaining inputs. Thus, this problem and its associated predicate require chip area of at least $\Omega(n)$ and have $AT^2 = \Omega(n^2)$. Furthermore, as discussed above the A^2T measure is a weaker measure. (Thompson [2] shows that $AT^2 = \Omega(n^2 \log^2 n)$ for this problem).

The binary sorting function $f_{BS}^{(n)} : \{0,1\}^n \to \{0,1\}^n$ puts an arbitrary binary n-tuple into ascending order. We show that $C_p^s(f_{BS}^{(n)}) = O(n)$ and that the product $AT = O(n)$ can be achieved for $A = O(\log n)$. Thus, binary sorting is clearly a very different problem from general sorting. Also, not only is A^2T a better measure for this problem, in fact, the measure AT and the computational inequality of Theorem 3 is even better.

The schema of Muller and Preparata [21] (also see [12] p. 73) has been used to realize $f_{BS}^{(n)}$ with a linear size (non-planar) circuit. This function is constructed from a circuit $f_c^{(n)} : \{0,1\}^n \to \{0,1\}^m$, $m = \lceil \log_2(n+1) \rceil$, which represents with m bits the number of 1's among the inputs. It is straightforward to show that this function has $C_p^s(f_c^{(n)}) = O(n)$. The next step is to apply the output of $f_c^{(n)}$ as input to a decoder. In Theorem 2 decoders with m inputs are shown to have planar circuit size linear in 2^m or in n. The output of the decoder contains a positional representation of the number of 1's in the input, which can easily be converted to the desired output with a linear number of elements in a planar circuit. (As a consequence of this construction, every 1-output Boolean symmetric function can be shown to have linear planar circuit size.)

A chip for this problem for which $AT = O(n)$, can be designed as follows. Group the n inputs into n/k groups of k each. Supply inputs to the chip in these groups and apply $f_c^{(k)}$ to each group and add to the previous count which is held in a register of length $O(\log n)$. After $T = n/k$ cycles have been executed, generate outputs in groups of k using a decoder and auxiliary logic, as described for $f_{BS}^{(n)}$. The chip has area $A = O(k + \log n)$, and $AT = O(n)$ if $k = \Omega(\log n)$. This chip algorithm is optimal to within a multiplicative factor.

Meyer and Paterson [12, p 43] have introduced a Boolean function $f_{MP}^{(n)} : \{0,1\}^n \rightarrow \{0,1\}$ that is contained in $P_{p,2^p}^{(n,1)}$ for $p = \Omega(n/\log n)$ and that has a a linear standard combinational complexity. It is not difficult to show that it also has a linear semelective planar circuit size. However, no such circuits are known which have all their inputs on the periphery of circuits. The function is defined on $n = kb$ variables which are grouped into k groups of b variables each. Here we can set $b = 2p+1$. On each group, the threshold function of threshold p+1 is defined to provide k output variables $a_0, a_1, \ldots, a_{k-1}$. These are used as an address and $f_{MP}^{(n)}(x_0, x_1, \ldots, x_{n-1}) = x_a$ where $a = a_{k-1}2^{k-1} + \ldots + a_1 2 + a_0$. Since the n+1 outputs of $f_{BS}^{(n)}$ are threshold functions and since x_a can be selected by a decoder with adress $(a_{k-1}, \ldots, a_0)$, a linear planar circuit can be designed.

From Theorem 5, $f_{MP}^{(n)}$ requires $A/\lambda^2 = \Omega(n/\log n)$. since $C_p^s(f_{MP}^{(n)}) = O(n)$, the A^2T measure is clearly non-optimal.

As the last topic of this section consider the question of contiguity of inputs and outputs. This is illustrated by the addition of two binary n-bit numbers. Assume that they are supplied as ordered contiguous bits to an adder whose boundary is a simple closed curve, as suggested in Figure 7. There is no ovelap of the bits in the two numbers. Then it is clear that node disjoint paths must exist between $x_{n/2-1}, \ldots, x_1, x_0$ and $z_{n/2-1}, \ldots, z_0$ and between $y_{n-1}, \ldots, y_{n/2}$ and $z_{n-1}, \ldots, z_{n/2}$. Thus, $(n/2)^2$ crossings must exist between these paths, given the perfectly general ordering of $\underline{x}$, $\underline{y}$ and $\underline{z}$ in Figure 7, so area proportional to n^2 is necessary. If contiguity is not required, a full adder chain for addition can be constructed that has linear area.

The same argument applies to integer multiplication under general assumptions about the representation of the integers. If the integers are represented in the standard binary number system, then it follows immediately from the lower bounds on AT^2 [3,4] that $A = \Omega(n^2)$ if $T = 1$ for n-bit integer multiplication. However, if integers are represented by exponents in their prime factor decomposition, multiplication can be done by adding exponents. In this case no semelective non-linear lower bound to AT^2 in the length of the representation of integers is known and in fact multiplication can be done with constant area if sufficient time is permitted. On the other hand, Yao [10] has shown that a quadratic lower bound holds for the function $x + y * z$. Consider, then, the case in which the product $c = ab$ is to be formed $0 \leq a,b \leq N-1$, the same system of (binary) representation is used for a, b, c, and each is given on a contiguous segment on the boundary of a simple closed curve. Suppose that as the curve is traversed, that the bits in the representations of a and b occur in the same order and that those of c occur in reverse order. (If those of c occur in the reverse order to those of a, say, the proof is also straightforward.) Each of a and b can be identified with c so we consider the "low order" k bits of each for the smallest value of k such that at least $\lfloor\sqrt{N}\rfloor$ different values are assumed by these k bits. The remaining bits must assume at least $\lfloor\sqrt{N}\rfloor/2$ values. Thus, at least $(\log_2 \lfloor\sqrt{N}\rfloor) - 1$ bits of information must flow from low order bits of a or b to c, and the same for the high order bits. Now apply the same argument used for binary integer addition.

It should be noted that it is standard practice to designer adders by overlapping two registers on a common bus, loading each individually, and then adding the two intermingled integers, thus avoiding quadratic sized area (but perhaps only delaying the inevitable). It is not known if a similar construction will work for integer multiplication with some representation of integers.

These simple observations are very important and reflect the fact that chips containing a variety of functional units may present certain wiring incompatibilities that may be much more costly in area than any requirements imposed by functionality considerations.

6 - CONCLUSION

Planar combinational complexity is shown to provide an intermediate step in the derivation of simultaneous lower bounds on chip area and time as measured by the number of cycles of execution. The planar circuits, the size of which is used to state inequalities, can be required to be semelective or not according to whether or not the VLSI algorithms are semelective. It should also be noted that if I/O is done on the boundary of a VLSI chip, the two types of planar circuits which are constructed to simulate the chip computation will also have I/O on their boundaries. Therefore, planar circuit size with input and output variables on the periphery, that is, for which each input and output node is on the same face of the plane, is a new constraint which needs to be explored. For example, the function $f_{MP}^{(n)}$ considered in the previous section has a linear semelective planar circuit size when this condition is not imposed but appears to require size $\Omega(n\log n)$ otherwise.

We also have noted that the gap between planar and non-planar circuit sizes is at most quadratic but that no functions for which the gap is this large are known. On a more technical level, the gap between the bounds on these two measures for most Boolean functions, as stated in Theorem 2, probably can be tightened.

We have seen that for all multi-output functions for which quadratic lower bounds on C_p^s exist, similar bounds exist for their associated predicates. We also have derived such bounds for powers and reciprocal powers of binary integers by reducing them to the shifting function. More refined reductions are possible so that the coefficients can be improved and so that the results can be made practicable.

Before closing, we note that Baudet [23] has extended a lower bound of $AT^2 = \Omega(n\log^2 n)$ of Johnson [22] for binary addition to $APT = \Omega(n\log^2 n)$ for pipelined chips. Also, Chazelle and Monier [16] have shown a lower bound of $AT^2 = \Omega(n^2)$ for this problem. In each case, the

bounds reflect a time T measured not in cycles but in physical propagation time. The Johnson and Baudet results use the fact that a circuit with n inputs and one output must have depth of at least $\lceil \log_2 n \rceil$. Chazelle and Monier further argue that with many semiconductor technologies this time may be $\Omega(\sqrt{n})$, or $\Omega(n)$ if the inputs lie on a convex boundary. In each case, the authors are combining the number of execution cycles with their length. All results in this paper are stated in terms of the number of cycles.

APPENDIX A

THEOREM A1 : *For any* $0 < \delta < 1$ *a fraction of at least* $1 - 2^{-\delta m 2^n}$ *of the functions* $f : \{0,1\}^n \to \{0,1\}^m$ *have*

$$C(f) \geq \frac{m2^n}{n} (1 - \delta - O(n))$$

if $\log_2 m = O(n)$.

PROOF : The bound is obtained using a standard counting argument. Without loss of generality consider optimal semelective circuits for functions. Such a circuit with c logic elements has c + n nodes, including those associated with inputs. The circuit has 2c edges on its c logic elements each of which could be attached to the outputs of at most (c+n-1) nodes to form at most $(c+n-1)^{2c}$ directed graphs. The inputs nodes contain unique labels but each logic node can carry one of the 16 labels for the sixteen logic elements on 2 inputs.

Thus, there are at most $M = (16)^c (c+n-1)^{2c}$ labeled, directed graphs and they compute at most M/c! different sets of c functions since there are c! different permutations of outputs of each graph in the ensemble thus created. Each such distinct graph computes $(c+n)^m$ sets of m functions $f_j : \{0,1\}^n \to \{0,1\}$, that is $(c+n)^m$ functions $f : \{0,1\}^n \to \{0,1\}^m$ can be associated with each graph. Consequently, since $c! \geq c^c e^{-c}$, at most $N(c) = (16e)^c (1+(n-1)/c)^c (c+n-1)^c (c+n)^m$ functions $f : \{0,1\}^n \to \{0,1\}^m$ can be computed by optimal semelective circuits with c logic elements. Using the inequality $(1+a/b)^b \leq e^a$, it follows that

$$N(c) \leq [(16e)(c+n-1)]^{c+n-1} (c+n)^m$$

Hence, the total number of functions $f : \{0,1\}^n \to \{0,1\}^m$ that can be realized by circuits with C or fewer elements is at most $(C+1)N(C)$ which is bounded above by

$$P = [(16e)\ (C+n+m)]^{(C+n+m)}$$

Now let C be such that $P = 2^{m2^n(1-\delta)}$. Then, since

$$y\log_2 y \geq A$$

implies

$$y \geq A/\log_2 A$$

for $A \geq 2$ we have that

$$C + n + m \geq \frac{m2^n(1-\delta)}{\log_2[(16e)m2^n(1-\delta)]}$$

which implies that

$$C \geq m\frac{2^n}{n}(1-\delta-O(n))$$

if $\log_2 m = O(n)$. Thus the fraction of the functions $f : \{0,1\}^n \to \{0,1\}^m$ that can be realized with this many or fewer logic elements is $P/2^{m2^n}$ which establishes the desired lower bound. □

7 - REFERENCES

[1] C.D. Thompson
"Area Time Complexity for VLSI",
Procs. 11th ACM Ann. Symp. Th. Comp., pp. 81-88, April 1979.

[2] C.D. Thompson
"A Complexity Theory for VLSI",
Ph.D. thesis, Dept. of computer Science, Carnegie-Mellon University, Pittsburgh, PA, August 1980.

[3] R.P. Brent and H.T. Kung
"The Area-Time Complexity of Binary Multiplication",
Report No. CMU-CS-79-136, Dept. of Comp. Sci., Carnegie-Mellon U., Pittsburgh, PA, July 1979. To appear in JACM.

[4] H. Abelson and P. Andreae
"Information Transfer and Area-Time Tradeoffs for VLSI Multiplication",
CACM 23, pp. 20-23, January 1980.

[5] J.E. Savage
"Area-Time Tradeoffs for Matrix Multiplication and Related Problems in VLSI Models",
To appear in Journal of Computer and System Science.

[6] H.T. Kung and C.E. Leiserson
"Algorithms for VLSI Processor Arrays",
pp. 271-292 in Introduction to VLSI Systems, ed. C. Mead and L. Conway, Addison-Wesley, Reading, MA, 1980.

[7] L.J. Guibas, H.T. Kung and C.D. Thompson
"Direct VLSI Implementation of Combinatorial Algorithms",
Proc. Conf. Very Large Scale Integration : Architecture, Design, Fabrication, California Institute of Technology, January 1979

[8] F. Preparata and J. Vuillemin
"Area-Time Optimal VLSI Networks for Multiplying Matrices"
Info. Proc. Letters, Vol. 11, No. 2, pp. 77-80, October 20, 1980.

[9] J.E. Vuillemin
"A Combinatorial Limit to the Computing Power of VLSI Circuits",
Proceedings of the 21st Annual Symposium on Foundations of Computer Science, pp. 294-300, October 1980.

[10] A.C. Yao
"The Entropic Limitations of VLSI Computations",
To appear in Procs. 13 th ACM ann. Symp. Th. Comp., 1981.

[11] R.J. Lipton and R.E. Tarjan
"Applications of a Planar Separator Theorem",
SIAM J. Comput, Vol 9, No. 3, August 1980.

[12] J.E. Savage
The Complexity of Computing, Wiley-Interscience, New-York, 1976.

[13] O.B. Lupanov
"A Method of Circuit Synthesis",
Izv. V.U.Z. Radiofiz., Vol. 1, No. 1, pp. 120-140.

[14] J.E. Savage
"Computational Work and Time on Finite Machines",
JACM, Vol. 19, No. 4, pp. 660-674, 1972.

[15] C.L. Seitz
"Self-Timed VLSI Systems",
Procs. Caltech Conference on VLSI, pp. 345-354, January 1979.

[16] B. Chazelle and L. Monier
"Towards More Realistic Models of Computation for VLSI",
Procs. 13 ACM Ann. Symposium Th. Comp., 1981.

[17] A.C. Yao
"Some Complexity Questions Related to Distributive Computing",
Procs. 11th ACM Ann. Symp. Th. Comp., pp. 209-213, April 1979.

[18] R.J. Lipton and R. Sedgewick
"Lower Bounds for VLSI",
To appear in Procs. 13 th ACM Ann. Symp. Th. Comp., 1981.

[19] G. Baudet
Personnal communication.

[20] C. Heinz
Personnal communication.

[21] D.E. Muller and F.P. Preparata
"Bounds to the Complexities of Networks for Sorting and Switching",
JACM, Vol. 22, No. 2, pp. 195-201, 1975.

[22] R.B. Johnson
"The Complexity of a VLSI Adder",
Vol. 11, No. 2, Info. Proc. Letters, pp. 92-93, October 20, 1980.

[23] G.M. Baudet
"On the Complexity of VLSI Circuits : A New Tool",
Preprint, INRIA, Rocquencourt, 1980.

[24] L.G. Valiant
personal communication.

[25] D. Yu. Grigoryev
"An Application of Separability and Independence Notions for Proving Lower Bounds on Circuit Complexity",
Notes of Scientific Seminars, Steklov Math. Inst. Vol. 60,
pp. 35-48, 1976.

[26] M. Tompa
"Time-Space Tradeoffs for Computing Functions, Using Connectivity Properties of Their Circuits",
Proc. 10th Ann. ACM Symp. Th. Comp., pp. 196-204, May 1978.

[27] J. E. Savage and S. Sewamy
"Space-Time Tradeoffs for Oblivious Integer Multiplication",
Lecture Notes in Computer Science, ed. H.M. Maurer, Springer-Verlag, Berlin, Heidelberg, New-York, pp. 498-508, July 1979.

[28] J. Ja'Ja'
personal communication.

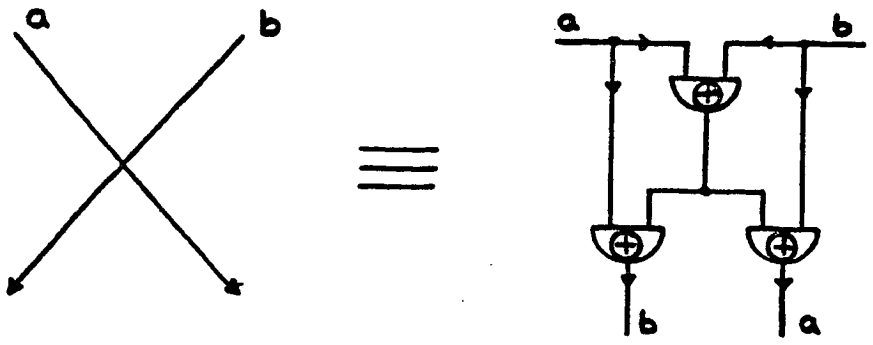

FIG. 1 REPLACEMENT OF CROSSING WIRES BY PLANAR EXCLUSIVE OR CIRCUIT

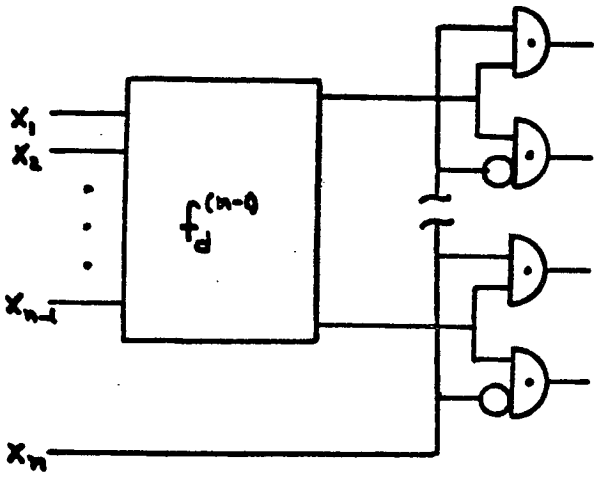

FIG. 2 RECURSIVE REALIZATION OF CIRCUIT FOR DECODER $f_d^{(n)}$

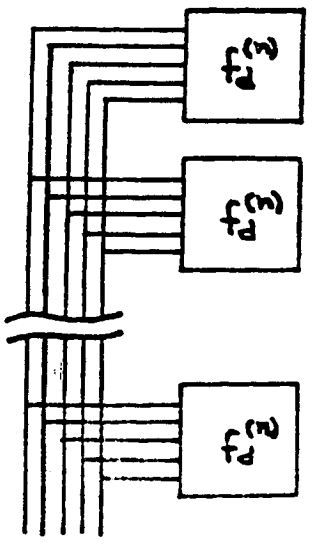

FIG. 3 BUS STRUCTURE FOR MULTIPLE DECODERS

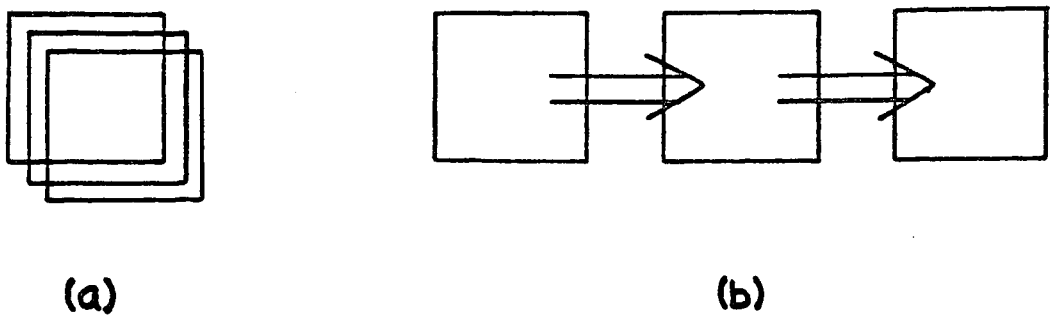

FIG. 4 SCHEMAS FOR CONSTRUCTING PLANAR CIRCUITS

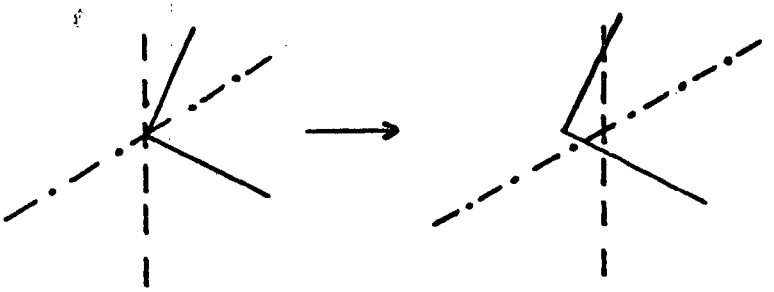

FIG. 5 DISPLACEMENT OF CROSSING WIRES

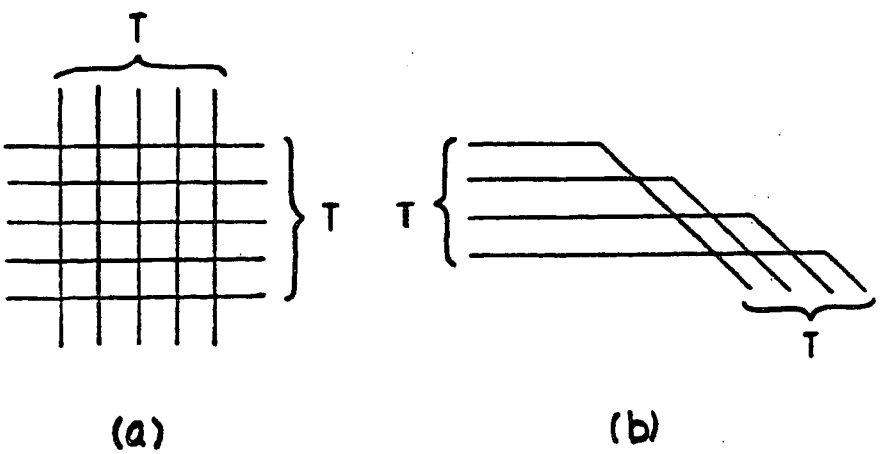

FIG. 6 CROSSINGS IN A STACK OF T DISPLACED CHIPS

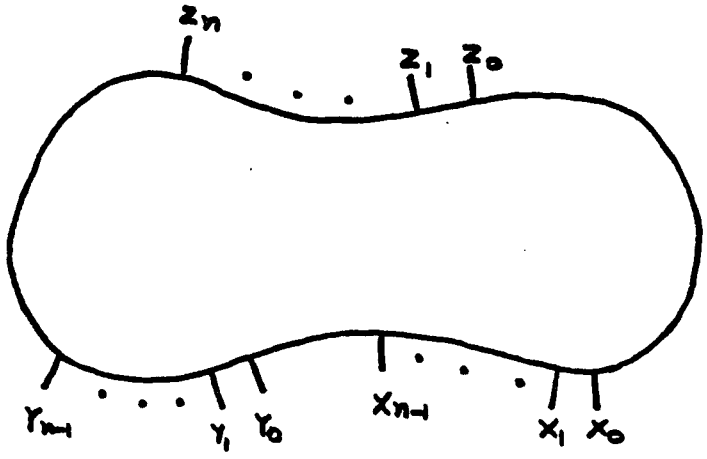

FIG. 7 · CONTIGUOUS INPUTS AND OUTPUTS

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique